DOCUMENT RESUME

ED 193 309 TM 800 624

AUTHOR Boldovici, John A.; Drucker, Eugene H.
TITLE The Negligible Effects of Process-Product Distinctions on Evaluation Quality.
PUB DATE Oct 79
NOTE 8p.; Paper presented at the Annual Meeting of the Military Testing Association (21st, San Diego, CA, October 15-19, 1979).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Evaluation Methods; *Formative Evaluation; *Measurement; Measurement Objectives; *Summative Evaluation

ABSTRACT

Sources of confusion about the distinction between process and product measurement are identified, and the effect of the distinction on evaluation quality is shown. Process measurement involves the measure of behavior, performance of a task which generates no products, an independent variable or treatment, use in diagnosing performance, and means. Product measurement involves the measure of a trace ("product") of behavior, performance of a task which generates products, a dependent variable or outcome, use in certifying performance, and ends. Four kinds of evaluation may result from various combinations of precision and imprecision in standards on the one hand, and measurement or estimation of behavior or products on the other. Evaluation improves with the extent to which it consists of comparing measured characteristics of behavior or products with well defined standards, and its quality does not depend on whether process (behavior) or product (trace of behavior) has been measured. (RI)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED193309

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
J. Boldovici

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

THE NEGLIGIBLE EFFECTS OF PROCESS-PRODUCT DISTINCTIONS ON EVALUATION QUALITY

John A. Boldovici
and
Eugene H. Drucker

Human Resources Research Organization (HumRRO)
Fort Knox, Kentucky 40121

TM 800624

Much has been written and said during the past few years about process and product measurement. The purposes of this paper are to identify sources of confusion about the process-product distinction, and more importantly, to show that the process-product distinction has little or no effect on evaluation quality.

The difference between process and product measurement obviously is not in the measure. One cannot tell whether 2.5 cm is a process or a product measure any more than one can tell whether 2.5 cm is a measure of the radius or the diameter of a circle. If not by inspection of a measure, how is the distinction to be made? It can be made along at least five dimensions. All writers are not in agreement as to which dimensions should be used (compare, for example, Osborn, 1973; with Army, 1977), and some writers switch from one dimension to another, without acknowledging the implications of having done so. (See, for example, Osborn, 1973; and Swezey and Pearlstein, 1975). The five dimensions along which the process-product distinction is made are shown in Table 1, and involve measuring:

1. By observation of behavior, as opposed to some trace ("product") of behavior. Guidance for developing Skill Qualification Tests (Army, 1977; Campbell, Ford, and Campbell, 1978) is unequivocal in distinguishing between process-scoring and product-scoring depending on whether,

   > ... the scorer observe[s] ... the task performance or ... some product produced during performance of the task (2, p. 4.7).

   The distinction seems reasonable and is easy to apply. Its adoption has not been widespread.

TABLE 1

FIVE WAYS TO DISTINGUISH BETWEEN
PROCESS AND PRODUCT MEASURES

Process If The Measure Is Of:	Product If The Measure Is Of:	References
Behavior	Trace ("Product") of Behavior	Army, 1977; Campbell, Ford, and Campbell, 1978
Performance of a Task Which Generates No Products	Performance of a Task Which Generates Products	Osborn, 1973; Swezey and Pearlstein, 1975
An Independent Variable or Treatment	A Dependent Variable or Outcome	Fitzpatrick, 1970
Use in Diagnosing Performance	Use in Certifying Performance	Osborn, 1973
Means	Ends	

2. Performance of tasks which do not generate "products" as opposed to tasks which do. Osborn (1973), and Swezey and Pearlstein (1975) referring to Osborn's work, distinguish between tasks in which the process is the product (close-order drill and diving, for example), and tasks in which the product always follows correct performance of the process (packing a parachute, for example). Osborn (1973) also notes, "Relatively few tasks are of the first type--those in which the product and the process are the same." Indeed, one might argue that no tasks are of the first type. Diving is no more the product of diving than parachute-packing is the product of parachute-packing. If a packed parachute is the product of packing a parachute, then a splash in the water or a wet diver must be the product of diving. The difference between the tasks is not that the product is the process in one and not in the other. Rather, one difference is in the concreteness or permanance of parachutes as compared to splashes--a difference with no compelling implications for measurement, because it is easily eliminated by the use of photography or other means of making permanent records. A more important difference is that one can, with a minimum of practice, inspect parachutes and make useful inferences about the quality of task performance, but cannot inspect splashes or wet divers and make very useful inferences about diving. This difference inheres, not in the tasks themselves, but in our ignorance about the relations between task outcomes and performance quality. Where little or no special knowledge is required to relate outcomes to performance, the performances typically are judged on the basis of outcomes. Footraces are examples. Where special knowledge is required (and absent), we relegate performance assessment to judges and critics. Evaluating diving, concerts, and gymnastics are examples.

3. Independent, as opposed to dependent, variables. Fitzpatrick (1970), in discussing measurement for program evaluation, notes that process measures describe the program (the independent variable) as it occurred, and outcome measures (the dependent variables) identify program effects.

4. For diagnosis, as opposed to certification. Osborn (1973), in describing the uses of process and product measurement, introduces a distinction based on test purpose:

> ... measures which focus on task outcomes (products) normally provide data relevant to the first purpose [certification], whereas measures of how tasks are carried out (process) pertain to the second [diagnosing instructional weaknesses] (p. 1).

This is a variant of another dimension for making the distinction; namely, a means-end dimension.

5. Means, as opposed to ends. Since processes and products are roughly analogous to means and ends, viewing all measures of means as process measurement, and all measures of ends as product measurement seems reasonable. The means-end dimension is useful because it implies the need to specify objectives: without objectives, one cannot determine whether means or ends have been measured.

Distinguishing between process and product measurement is not difficult if the measured event or product falls at the same end of all five dimensions. If, for example, a measure is made by direct observation of performance of a task which has products from which inferences cannot be made about performance quality, and the measure is of an independent variable and used diagnostically, and the task is a means to an end, then one is clearly dealing with a "process" measure. Confusion arises, however, when a measured event or product falls near opposite ends of any two dimensions. Suppose for example that measures taken from a sight photograph were used diagnostically. The SQT experts would say that product measurement had taken place, because observation was not made of behavior. The measures would, however, qualify as process if the diagnosis-qualification dimension were used. Whether the measures were of means or ends would depend on the stated objective: means (process) if the objective were to hit a target, ends (product) if to lay cross-hairs on an aiming point.

Despite apparent confusion about appropriate dimensions for distinguishing between process and product measurement, the distinction may be useful--albeit indirectly. The distinction reflects a concern with objective, as opposed to subjective evaluation. If evaluation can be defined as making judgments based on comparisons between expected or desired characteristics of behavior or a product of interest, and observed characteristics of the same behavior or product, then it seems to follow that the more precisely stated the expectations and the more objective the observations, the better the

evaluation. This line of thinking underlies current emphases in criterion-referenced testing: stating standards (expectations) on the one hand, and objectively measuring performance (or "products" of the performance) on the other. Notice, however, that standards can be well defined, loosely stated, or not defined or stated at all. Examples near the extremes are, "At least one target hit with two rounds," and "neutralizes targets at various ranges." Similarly, behavior or product characteristics can be measured precisely, measured imprecisely, or measured not at all. Examples near the extremes are measuring length with a ruler and estimating casualties. Evaluation improves with the extent to which it consists of comparing measured characteristics of behavior or products with well defined standards.

Figures 1 and 2 present examples of kinds of evaluations which result from various combinations of precision and imprecision in standards on the one hand, and measurement or estimation of behavior or products on the other. The upper left cell in both figures is the most desirable from the standpoint of good evaluation. Here measured characteristics are compared with well defined standards. When a physician says, "Your pulse rate is high," we have confidence in this evaluation because (a) the physician has measured our pulse rate objectively, (b) the standard is clearly defined, and (c) the rules for making a judgment of high or low are obvious. For the same reasons we have confidence in evaluations of tensile strength (Figure 2, upper left cell.)

The upper right cells in both figures inspire less confidence, but retain the virtue of well defined standards. Estimates rather than measures of behavior or product characteristics are used here, probably for expediency, and could, with added cost, be replaced by measurement: A photograph could be made of the pass receiver and the boundary line (Figure 1), and instruments could be used to measure front-end play in the car (Figure 2). Comparing the results of the photographs or of the instrument readings with available standards for "out-of-boundness" or the need for front-end work would increase our confidence in the evaluation.

The bottom cells of the two figures are of interest mainly as areas to be avoided by evaluators. In the lower left cells the evaluator makes precise measures of the behavior or product of interest, and compares these with a standard which is loosely or not defined. "Evaluations" of mental health and of stereo equipment, for example, involve very precise measurement of many behavior and product characteristics whose relevance to a behavior or product standard cannot be known because the standard itself is not known, and the rules for making the comparisons are nearly as numerous as the evaluators.

		BEHAVIOR CHARACTERISTICS	
		Measured	Estimated
BEHAVIOR STANDARD	Well defined	Comparing observed with normal heart rate.	Referee calling a pass receiver out of bounds.
	Loosely or not defined	Evaluating a patient's mental health based on results of diagnostic tests.	Judging ice-skating, diving, dancing.

Figure 1. Four kinds of process evaluation.

		PRODUCT CHARACTERISTICS	
		Measured	Estimated
PRODUCT STANDARD	Well defined	Comparing measured tensile strength with design specifications.	Deciding whether a car needs front end work by shaking its wheels.
	Loosely or not defined	Consumer-oriented tests of stereo equipment, TV, washing machines, etc.	Judging a painting or a novel.

Figure 2. Four kinds of product evaluation.

In the last (bottom right) cells of both figures the evaluator is comparing the results of no measurement with a nonexistent or ill-defined standard. This is the realm of opinion and sophistry.

Perhaps the most interesting aspect of Figures 1 and 2, however, is that good evaluation is not the result of whether process (behavior) or product (trace of behavior) has been measured. It is rather a function of:

1. Precision in stating standards.

2. Objectivity in measuring the product or behavior characteristics.

3. Veridicality of rules for comparing measured characteristics and standards.

This is so regardless of whether processes or products are being evaluated.

END

DEPT. OF HEW

NAT'L INSTITUTE OF EDUCATION

ERIC

DATE FILMED

FEB. 11 1981